



HAL
open science

ELAL: An Emotion Lexicon for the Analysis of Alsatian Theatre Plays

Delphine Bernhard, Pablo Ruiz Fabo

► **To cite this version:**

Delphine Bernhard, Pablo Ruiz Fabo. ELAL: An Emotion Lexicon for the Analysis of Alsatian Theatre Plays. Conference on Language Resources and Evaluation (LREC 2022), Jun 2022, Marseille, France. pp.5001-5010. hal-03655148

HAL Id: hal-03655148

<https://hal.science/hal-03655148>

Submitted on 16 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

ELAL: An Emotion Lexicon for the Analysis of Alsatian Theatre Plays

Delphine Bernhard, Pablo Ruiz Fabo

Université de Strasbourg, LiLPa UR 1339

F-67000 Strasbourg, France

{dbernhard, ruizfabo}@unistra.fr

Abstract

In this work, we present a novel and manually corrected emotion lexicon for the Alsatian dialects, including graphical variants of Alsatian lexical items. These High German dialects are spoken in the North-East of France. They are used mainly orally, and thus lack a stable and consensual spelling convention. There has nevertheless been a continuous literary production since the middle of the 17th century and, in particular, theatre plays. A large sample of Alsatian theatre plays is currently being encoded according to the Text Encoding Initiative (TEI) Guidelines. The emotion lexicon will be used to perform automatic emotion analysis in this corpus of theatre plays. We used a graph-based approach to deriving emotion scores and translations, relying only on bilingual lexicons, cognates and spelling variants. The source lexicons for emotion scores are the NRC Valence Arousal and Dominance and NRC Emotion Intensity lexicons.

Keywords: emotion, lexicon, Alsatian, theatre

1. Introduction

Automatic emotion analysis for written corpora of theatre plays has been little investigated to date (see Section 2.3). This is nevertheless a promising avenue, since previous work based on novels has shown that the characters’ direct speech contains salient emotional features (Henny-Krahmer, 2018). There are two ways in which emotions can be expressed in theatre plays: (i) Directly in characters’ speech, through emotion-bearing words; (ii) Indirectly in stage directions, which indicate how a character should behave or look like on stage. For instance, Kim and Klinger (2019) have analysed a corpus of fan fiction to show how emotions are also expressed non-verbally, e.g., facial expressions, gaze, voice characteristics, gestures, body postures. Sentiment lexicons are an important resource for automatic emotion analysis. While expressions in sentiment lexicons should be interpreted with respect to contextual cues (negations, long distance dependencies, ambiguity), lexicon-based approaches are nevertheless easier to implement in a low-resource setting, where annotated data amenable to supervised learning are notably lacking. In this work, we focus on the Alsatian dialects, High German dialects spoken in the North-East of France. These dialects are used mainly orally, and thus lack a stable and consensual spelling convention. There has nevertheless been a continuous literary production since the middle of the 17th century and, in particular, theatre plays. Recently, image-mode digital versions for a large sample of theatre plays, representative of the tradition, were created by the BNU.¹ Within the MeThAl project (Ruiz Fabo et al., 2021),²

¹The BNU is Strasbourg’s University and National Library. The plays are at <https://numistral.fr/fr/theatre-alsacien>, accessible via the *Découvrir* (i.e. *Discover*) link.

²<https://methal.pages.unistra.fr>

we are currently performing Optical Character Recognition (OCR) on this corpus and encoding it according to the Text Encoding Initiative (TEI) Guidelines (TEI Consortium, 2021). The overall project goal is enabling large-scale computationally supported analyses of the Alsatian dramatic tradition, which would not be possible without the TEI-encoded corpus under creation, as well as starting off such analyses, which have not yet been carried out for Alsatian theatre.

In this paper, we present a method to build multilingual sentiment lexicons, relying on existing sentiment lexicons originally built for English and including translations into French and German.

The main contributions of the paper are:

- A novel and manually corrected emotion lexicon for the Alsatian dialects, including graphical variants of Alsatian lexical items;
- A graph-based approach to deriving emotion scores and translations, relying only on bilingual lexicons, cognates and spelling variants.

The article is structured as follows. We first address the state of the art in Section 2. Section 3 is devoted to the description of our method. Results are analysed in Section 4. Finally, we discuss information relevant to future uses of the lexicon in Section 5.

2. State of the Art

2.1. Sentiments and Emotions

Kim and Klinger (2021) use the following definitions, relying on previous work in the domain:

- **sentiment**: “a *positive* or *negative* feeling underlying the opinion”;
- **emotion**: a feeling that “involve[s] a set of expressive, behavioral, physiological, and phenomenological features”.

They stress that the identification of emotions is more complex than that of sentiments (binary polarities):

there are more emotion classes and they are more difficult to distinguish from one another. Emotions can be subdivided into several sets, as quoted by Kim and Klinger (2021):

- Ekman's emotions: *anger, fear, joy, sadness, surprise, disgust.*
- Plutchik's eight basic bipolar emotions: *joy vs. sadness, anger vs. fear, trust vs. disgust, surprise vs. anticipation.*

In addition, the valence-arousal-dominance model attempts to represent emotions not as categories, but in three continuous dimensions (Bradley and Lang, 1994), quoted by Kim and Klinger (2021).

Emotion categories, as well as the three dimensional valence-arousal-dominance model, are the foundations for many sentiment or emotion lexicons.

2.2. Lexicons for Sentiment and Emotion Analysis

Several lexicons for sentiment and emotion analysis have been developed for the English language: the General Inquirer's lexicons (Stone et al., 1966), the MPQA subjectivity lexicon (Wilson et al., 2005) or SentiWordNet (Esuli and Sebastiani, 2006), among others. In particular, lexicons for emotion analysis were manually created at the National Research Council Canada. These lexicons include:

- the NRC Emotion Intensity Lexicon (NRC-EIL), which details scores of intensity for English words with respect to Plutchik's eight basic emotions (Mohammad, 2018b). The words included in this lexicon do not necessarily denote an emotion directly, but can connote an emotion, with a specific degree of intensity. Words have been manually annotated using the *best-worst scaling* annotation scheme which consists in performing comparative annotations.
- the NRC Valence, Arousal and Dominance Lexicon (NRC-VAD), which details valence (positive-negative), arousal (active-passive) and dominance (dominant-submissive) scores for English words (Mohammad, 2018a). This lexicon was also created using best-worst scaling.

While early work was mostly targeted at English, there have been more recent attempts to build emotion lexicons for other languages from scratch, or by projecting existing lexicons to other languages.

A first basic approach consists in using a machine translation tool to translate English words to target languages, while keeping the original emotion scores obtained for English. The NRC-EIL and NRC-VAD have been translated to more than 100 languages using Google Translate. On the download page, it is argued that “[d]espite some cultural differences, it has been shown that a majority of affective norms are stable across languages.”³

³<http://saifmohammad.com/WebPages/AffectIntensity.htm>

More recent approaches rely on cross-lingual word embeddings. Ramachandran and de Melo (2020) present a method to address low resource languages, which do not have machine machine translation systems. They use cross-lingual word embeddings encompassing both the source language, which possesses emotion ratings, and the target language. Emotion ratings are then induced by finding the most similar source language words in the cross-lingual word vectors. For their experiments, they use the text of the Bible for about 350 different languages, English being the source and resource-rich language (NRC-EIL). Unfortunately, the results obtained for the lesser resourced languages, with less than 10K verses in the Bible are not satisfactory, and this is confirmed by the results obtained on the task of predicting the emotion of sentences.

Buechel et al. (2016) describe a method to induce VAD lexicons for historical 18th and 19th century German. The proximity of historical variants to their contemporary counterpart is determined using word embeddings, trained on a large corpus of historical content (*Deutsches Textarchiv*) and the affective scores are determined using a contemporary seed lexicon. The expanded lexicon is then used to measure textual emotion in historical texts. They show that different text genres are characterised by different VAD patterns, and that chronological shifts can also be revealed.

More recently, Buechel et al. (2020) have proposed an approach to obtain emotion lexicons for 91 languages, relying on bilingual word translation models and target language word embeddings. For word translation from English the target language, the Google Cloud Translation API is used, and *fastText* embeddings trained on Wikipedia and CommonCrawl (Grave et al., 2018) are used for the target languages. The evaluation results show good performance for about two thirds of the target languages, but lower results for the last third, which includes languages such as Yoruba, Corsican, Cebuano, or Somali, among others. The authors hypothesise that the drop in performance is due to the lower quality of the machine translation and of the word embeddings.

Based on the results of the previously described approaches, we decided not to use word embeddings: we only have small corpora at our disposal and the Alsatian dialects are characterised by a large amount of graphical variation. Instead, we decided to rely on the resources at our disposal, bilingual French-Alsatian lexicons, and the proximity between Alsatian and German.

2.3. Emotion and Sentiment Analysis for Theatre Plays

There is to date comparatively little work on automatic sentiment analysis for drama, in contrast to, say, online reviews or social media.

Mohammad (2011) analyses the distribution of emotion words in plays by William Shakespeare, measuring their relative salience in pairs of plays and measuring their evolution along the plays' timeline.

Nalisnick and Baird (2013a) describe a method to analyse the emotion dynamics between characters in plays by William Shakespeare. They use a sentiment lexicon and sum the valence values for continuous speech units, using this value to assess the sentiment directed to the preceding speaking character. This analysis can be used to detect changes in sentiments between character pairs in the course of the play. In another article, Nalisnick and Baird (2013b) additionally compared the average word valence in tragedies and comedies, to assess whether the two genres could be distinguished from one another and detect outliers. They also tried to combine sentiment and network analysis to detect factions between characters.

Schmidt and Burghardt (2018) and Schmidt et al. (2018) detail methods to perform sentiment analysis on German plays from the 18th century written by Gotthold Ephraim Lessing. They performed a manual annotation on 200 speeches into more detailed *differentiated polarity* categories and *binary polarity*. They compared several lexicon-based methods to perform automatic classification: use of different sentiment lexicons, extension of the lexicons with historical linguistic variants, use of stop words lists, lemmatisation, effect of lowercasing. In particular, they show that the extension of the lexicons with historical linguistic variants yields the largest increase in performance.

Schmidt et al. (2021) perform emotion classification in German plays between 1650 and 1815. A taxonomy was developed with 6 main emotions, 13 sub-emotions and two overall polarity classes (positive/negative). Over 13,000 manual annotations were created as a training corpus. Based on this corpus, different classifiers were developed, both with traditional and transformer-based methods; the latter were pretrained on historical and contemporary varieties. Best results were achieved with large transformer models pretrained on contemporary German.

While Klinger et al. (2020) focus on fiction, their work is nevertheless also relevant for theatre plays since they use a text's structure to analyse emotions; theatre plays usually have a clearly marked structure: division into acts and scenes, stage directions, indication of the name of the characters. In particular, Klinger et al. analyse emotional character interactions in English fan-fiction short stories, by annotating the characters that feel an emotion and also the character provoking that emotion, if available. The goal is to extract emotional character networks automatically.

In brief, emotion lexicons have been applied in earlier work on the analysis of theatre plays, while more recent approaches used supervised training based on annotated data. For the time being, we chose to focus on lexicon-based approaches and in particular the extension of lexicons with spelling variants, which have been shown to be beneficial (Schmidt et al., 2018).

3. Methodology

3.1. Source Lexicons

Emotion lexicons We use the French and German translation of NRC-VAD, a resource of 20,007 words with VAD ratings ranging from 0 to 1 (Mohammad, S., 2018), and NRC-EIL, which contains 9,921 word-emotion pairs with scores ranging from 0 to 1 (Mohammad, S., 2020). In these lexicons, the original English word is given, along with its translation into French or German. We chose French and German because we have bilingual Alsatian-French lexicons at our disposal, and word forms in the Alsatian dialects are often very close to their standard German equivalent. For evaluation purposes, we also use the French FEEL emotion lexicon (Abdaoui et al., 2017), which is a French version of the English NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013) which has been manually reviewed by a professional translator (Abdaoui, 2017).

Alsatian-French bilingual lexicons Bilingual lexicons were extracted from different sources: online dictionaries, lemmas and their translations into French from an annotated corpus (Bernhard, 2019) and a published multilingual dictionary (Adolf, 2006). The final lexicon contains 73,179 French-Alsatian pairs.

Theatre plays vocabulary We extracted the vocabulary from 68 Alsatian plays which have been processed by OCR and manually corrected in the context of the MeThAl projet.⁴ This vocabulary contains 58,631 words.

3.2. Sentiment Lexicons as Networks

Our goal is to obtain an emotion lexicon for the Alsatian dialects, encompassing a large array of spelling variants found in the corpus of theatre plays. To do so, we need to be able to detect spelling variants and to relate Alsatian word forms to French and German words in the original NRC VAD and EIL lexicons. Several methods are used (see Figure 1 for an illustration):

- Direct translations from Alsatian to French are found in the bilingual French-Alsatian lexicons (e.g. the translation of Alsatian *Ardebewa* to French *tremblement de terre*). However, this bilingual lexicon is quite small and covers only 60.43% of the entries in the French NRC VAD lexicon, 63.81% of the French NRC EIL lexicon and 60.33% of the French FEEL lexicon.
- We also check whether there are Alsatian spelling variants (e.g. Alsatian *Arbeewa* and Alsatian *Ardbewa*) or closely related Alsatian and German word forms (e.g. Alsatian *Erdbewe* and German *Erdbeben*). Two forms are considered to be related if:

⁴The plays' text can be browsed at <https://methal.eu/ui/> and a TEI-encoded subset is at <https://git.unistra.fr/methal/methal-sources> (Ruiz Fabo, 2022); see Ruiz Fabo et al. (2020) for the encoding workflow.

- They have one identical Double Metaphone key (Philips, 2000; Bernhard, 2014). Double Metaphone transforms each input string into one (possibly two, to account for ambiguities) keys which are identical for words whose pronunciations are similar, e.g. “Ard-beewa” and “Erdbewe” both have the keys ‘ARTPF’ and ‘ARTPV’.
- The string similarity measure given by the Python `diffliplib`⁵ module exceeds a given threshold t . The similarity measure is applied to normalised word forms, with all accents stripped. The final threshold of $t = 0.75$ was selected by comparing the number of valid Alsatian word forms manually selected for inclusion the lexicon, for a subset of 20 English terms from the NRC lexicons. We measured the precision, recall and f-measure for threshold values ranging between 0.6 and 0.9, with a step of 0.05. The best trade-off between precision and recall was found for a string similarity threshold of 0.75.

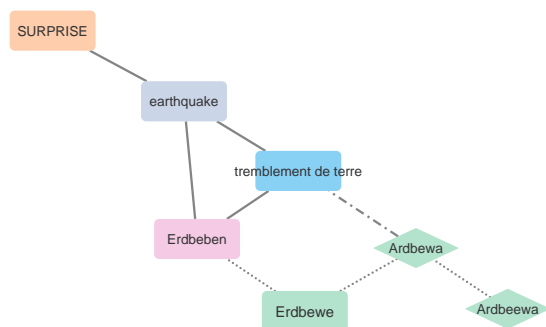


Figure 1: Visualisation of the emotion network in Cytoscape. The color of the nodes depends on their type: orange for emotions (*SURPRISE*), grey for English terms (*earthquake*), pink for German terms (*Erdbeben*), blue for French terms (*tremblement de terre*), green for Alsatian terms (*Erdbewe*, *Ardbewa*, *Ardbeewa*). Alsatian terms in rounded rectangles come from the corpus of theatre plays. Terms in diamonds come from the bilingual lexicons. The line used for the edges represents the source of the associations: solid line: NRC emotion lexicons; dash-dotted: French-Alsatian lexicons; dotted lines: graphical similarity (these are the most error-prone associations, since different lexical items can have closely related forms).

As a result, all the lexicons are aggregated and represented as a large network whose edges correspond to associations between words and emotions on the one hand, and between word forms in the four target languages on the other hand (translations and spelling variants). We use the `NetworkX` Python package (Hagberg et al., 2008) to represent our lexicons as net-

works. In order to ease the manual annotation and correction of these graphs, we output smaller sub-graphs of the whole graph, centred on a given emotion (based on the idea that it would be easier to correct together words related to the same emotion).

Such a representation has several advantages over a flat tabular format:

- These networks can be visualised and modified using network visualisation software; in this project, we used Cytoscape (Shannon et al., 2003).⁶
- Different types of relevant information can be highlighted thanks to styling (colours, shapes).
- Information coming from different sources can be combined in a single data structure.

3.3. Manual Correction

Even if we use bilingual French-Alsatian lexicons and set a threshold to limit the number of incorrect Alsatian translations and variants linked to German and French words, manual correction is nevertheless needed. We have to correct errors in the automatic identification of graphical variants for Alsatian words, errors found in the translations from the original NRC lexicons, as well as potential cases of semantic drifts due to translating from English to Alsatian through French or German.

The networks were manually corrected using the Cytoscape tool. In case of doubt, the corpus of plays as well as the bilingual Alsatian-French lexicons were consulted. The manual correction had the following objectives:

- Marking incorrect nodes as non valid. Figure 2 displays such an error. Here, the association between *Intrigen* (German) and *Intriguen* (Alsatian) is correct, but the terms *untergehn* and *undergehn* (cf. German *untergehen* – go down) have a different meaning and should be considered as non valid in this context. Note that French and German words could also be annotated as non valid, due to potential errors in the translations of the NRC lexicons.
- Adding missing edges between graphical variants of Alsatian words. Figure 3 displays an example where edges should be added between *jara* and *jare/jaare*, which are all graphical variants of the same lexical item.

The manual correction was performed by a linguistics master’s student, who is a native Alsatian speaker, during a paid internship. 1,199 sub-graphs, totalling 113,365 nodes and 368,868 edges, were corrected over a time period of 20 days (7 hours a day) amounting to about 140 hours of work. Due to this limited workforce, all the initial graphs could not be manually corrected. The number of nodes and edges might seem very large, but they are due to the spelling variation

⁵<https://docs.python.org/3/library/diffliplib.html>

⁶<https://cytoscape.org/>

als	fr	de	en	V	A	D	ang.	ant.	dis.	fea.	joy	sad.	sur.	tru.
Ardbeewa, Erdbewe, Ardbewa	tremblement de terre	Erdbeben, Beben	earthquake, quake	0.09	0.89	0.76	0.39			0.81		0.73	0.81	
ewerziigt, Überzeugt, iwerziigt, überzeugt	persuadé, convaincu, imbu	überzeugt	convinced	0.68	0.47	0.57								0.58
iwerziiga, ewerziiga, eww- erziiga, iwerzeiige, ewerziiga, ...	persuader, inciter, convaincre	überzeugen	convince, persuade	0.65	0.56	0.71		0.37						0.51
Schwindler, Schwëndler, Schwindler, Schwendler, Schwindlere, ...	escroc, faiseur, menteur	Schwindler, Lügner	escrow, swindler, crook, shys- ter, faker, liar	0.19	0.64	0.50			0.54					
nützlich, nüetzlich, nützlich, netzlig, nützlich, ...	utile	hilfreich, nützlich	helpful, usefully	0.74	0.41	0.72					0.42			0.60

Table 1: Example entries after applying personalised Page Rank.

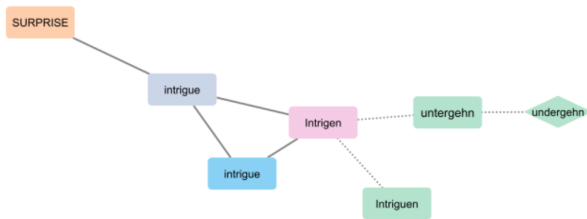


Figure 2: Incorrect associations in emotion networks – associations to *untergehn* and *undergehn* are incorrect.

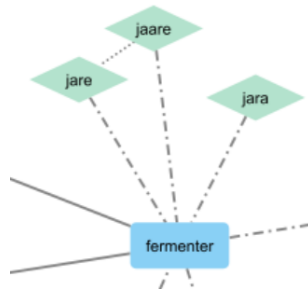


Figure 3: Missing associations in emotion networks – here *jara* should be associated to *jaare/jare*.

in the Alsatian dialects, leading to many observed potential variants which have to be checked in order to validate the translations.

3.4. Merging Corrected Networks

As a result of our manual annotation methodology, focusing on a single target emotion, a subset of words was present in several different subgraphs. For instance, the word “strength” is associated both to JOY and TRUST, and therefore this word and all its translations into French, German and Alsatian were annotated twice. We merge all these concurrent annotations. Nodes with diverging annotations are explicitly marked in the resulting merged networks to facilitate further checking.

After merging, intra-annotator agreement was computed, based on the English, French, German or Alsatian words which were annotated several times by the annotator. We consider an annotation to be consistent if the word was consistently annotated as valid or invalid. On the contrary, an annotation is considered as inconsistent if the annotation changes from valid to non-valid or from non-valid to valid in the course of the annotations (only the first change is considered and only words which were annotated at least twice are considered). The intra-annotator agreement measured with Cohen’s Kappa is equal to 0.80, with a percentage agreement of 91%.

3.5. Generation of the Final Lexicons

Finally, we generate the final lexicons. We first cluster together Alsatian words which are considered as graphical variants of a single lexical item. We then apply the personalised PageRank (Page et al., 1999) algorithm to

the graphs,⁷ starting from the Alsatian cluster nodes. The final emotion scores for a given cluster of Alsatian nodes is the average of the VAD and Emotion Intensity scores for English words⁸ in the graphs, weighted by their PageRank values.

Table 1 gives some examples of clusters of Alsatian word forms, their translations,⁹ and their scores. As can be seen, each cluster may be associated with several near-synonymous translations into English, French and German.

4. Analysis of the Results

The analyses and graphics were done using the following Python libraries: `pandas` (Reback et al., 2021), `matplotlib` (Hunter, 2007), `seaborn` (Waskom, 2021) and `scipy` (Virtanen et al., 2020).

4.1. Composition of the Final Lexicon

In the rest of the paper, we refer to the resource created as ELAL (Emotion Lexicon for ALSatian). Table 2 details the statistics for ELAL, including statistics for the subset of forms found in the theatre plays. Entries correspond to clusters of Alsatian variants, as exemplified by the rows in Table 1.

	Total	Theatre subset
Entries	5,932	3,273
Alsatian forms	22,923	11,920
Variants per entry	3.86	3.64
To check	6%	7%

Table 2: Statistics for ELAL.

4.2. Comparison with other Emotion Lexicons

4.2.1. Correlations with NRC and FEEL

In order to assess the quality of the final lexicon, we measure the standard correlations (Pearson’s r) between the emotion scores in ELAL and the NRC and FEEL lexicons (see Figure 4). We do not take into account vocabulary items which were marked as requiring further checking after manual annotations. When scores are missing, they are filled with 0.0.

The correlations with the NRC lexicons are consistently superior to 0.6, indicating a positive relationship between the scores, except for “anticipation” scores when compared to the NRC English lexicon. Correlations with the FEEL lexicon are lower, but scores in the FEEL lexicon are binary (0 or 1) while in ELAL and

⁷We use the implementation of PageRank provided in `NetworkX`.

⁸There are no specific scores for French and German words in the original NRC lexicons, since they only provide translations of English words.

⁹We only keep translations with a shortest path length inferior or equal to 2.

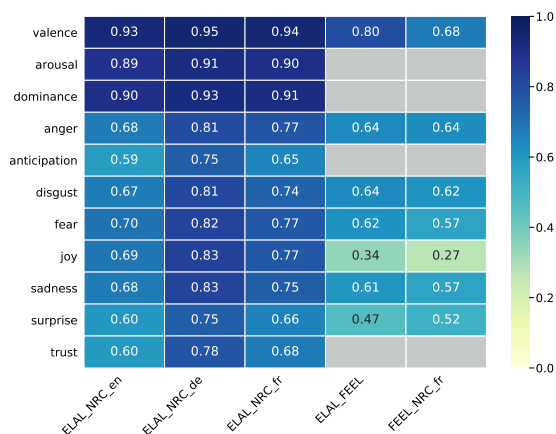


Figure 4: Correlations between the Alsatian (ELAL) the NRC (English, German and French) and the French FEEL lexicons.

NRC scores are continuous and range between 0 and 1. The same tendency for lower correlations is observed when comparing the French NRC lexicons to FEEL (last column in Figure 4). The correlations between ELAL and FEEL are generally superior to those between NRC and FEEL, except for the “surprise” emotion.

4.2.2. Coverage

We measure the coverage of ELAL with respect to the NRC and FEEL lexicons (see Figure 5). There is clearly room for improvement here: we saw in Section 3.2 that the coverage of our bilingual lexicon was about 60%. Since not all data could be manually corrected, this is further reduced to about 22 to 30% of the lexical items available in the NRC and FEEL lexicons.

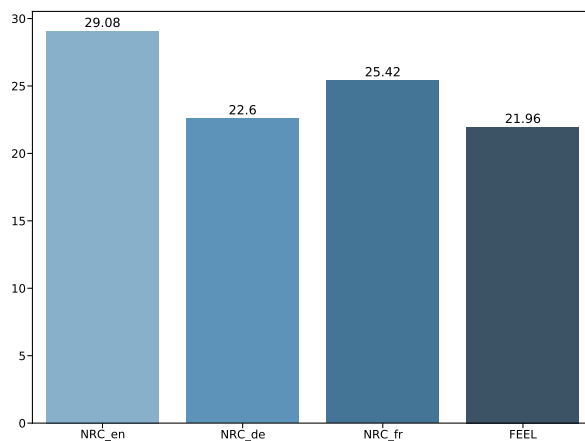


Figure 5: Coverage of the English, German and French NRC and the French FEEL lexicons.

4.2.3. Distribution of Scores

We compare the distribution of scores across emotion dimensions for the Alsatian and the original English

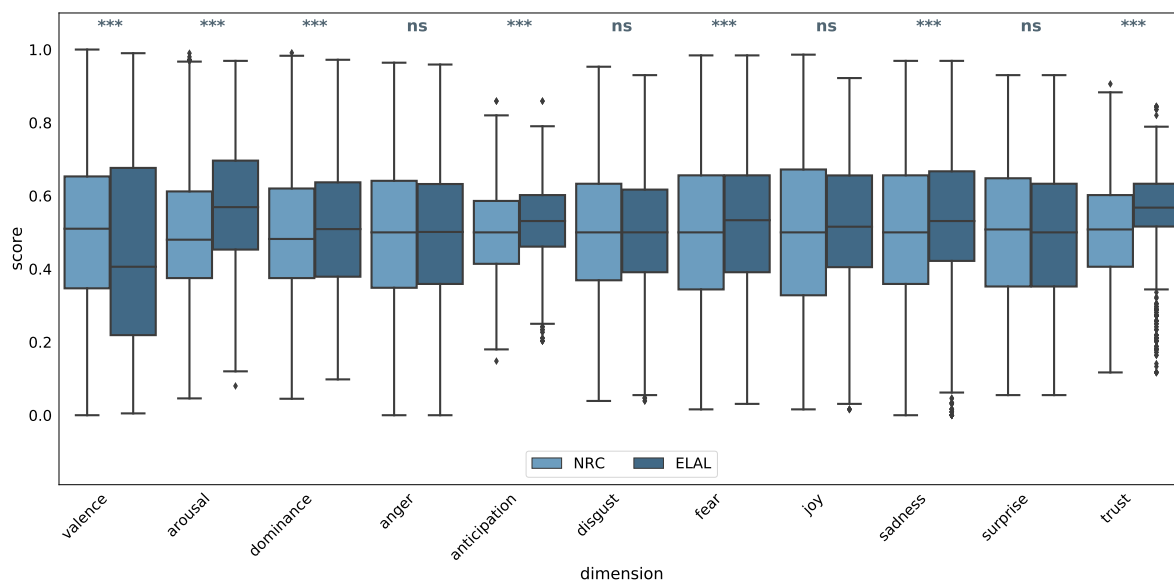


Figure 6: Box plot comparing the distribution of scores across the emotion dimensions, for the Alsatian (ELAL) and original English NRC lexicon. Statistically significant differences according to the Mann-Whitney U test ($p < 0.001$) are marked with ‘***’. Non statistically significant differences are marked with ‘ns’.

NRC lexicon (Figure 6). The differences are analysed using the non-parametric Mann-Whitney U test. They are all statistically significant ($p < 0.001$) except for four emotions: anger, disgust, joy and surprise. The categories for which the differences are not statistically significant also correspond to those for which a larger proportion of data was manually corrected. The differences for the other categories could be accounted for by the fact that our Alsatian lexicon only represents a fraction of the whole NRC lexicon, where the distribution of emotions is different from the distribution in the source lexicons. We can hypothesise that the differences would be lower had we manually annotated a larger proportion of the initial data.

4.2.4. Average Similarity

In order to perform a finer grained analysis at the level of single entries, we represent each entry as a vector using the VAD and emotion scores (only valence and six emotions for FEEL). We then compute the average of the cosine between this vector and the corresponding vectors in the NRC and FEEL lexicons for the translations into English and French of each entry. For the FEEL lexicon, the binary valence scores are scaled to the same range as the valence scores in ELAL, to prevent null vectors for FEEL entries.

Figure 7 displays the distributions of the cosine similarities with respect to the NRC and FEEL lexicons. These distributions are consistent with the observations made in Section 4.2.1: ELAL is closest to the NRC lexicons, with a majority of cosine values superior to 0.9. The cosine similarities with the FEEL lexicon are more dispersed: while they tend to be more dense towards highest similarities, 3.8% of the cosine similarities are

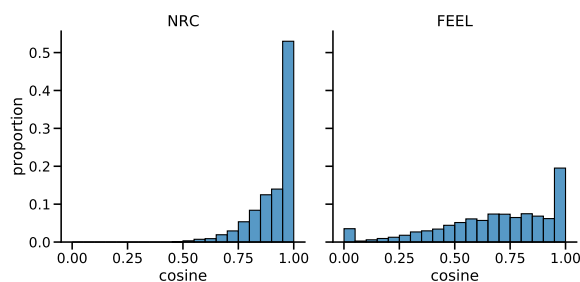


Figure 7: Distribution of cosine similarities with the NRC and FEEL lexicons.

inferior to 0.1, which might point at some potential errors in ELAL. Since cosine similarities are available for each entry, they can be used to filter the lexicon and keep only highest ranking entries, or detect discrepancies in the similarities with NRC and FEEL.

4.3. Emotions in the Theatre Plays

We briefly describe the usage of words from the emotion lexicon in the theatre plays. Figure 8 displays the distribution of the total relative frequency of words from the emotion lexicon in the plays, by decade. This figure shows that all plays are covered to a certain extent. Outliers are visible for decades 1920 and 1930:

- The play *A gelungener Patient*,¹⁰ published in 1931, has more emotion words, with a relative frequency of 12.60. Interestingly, the most frequent of them is the word "Doktr" (doctor), which is also

¹⁰<https://methal.eu/ui/text/html/huck-a-gelungener-patient/>

the name of one of the two characters.¹¹

- The play *Die Greifensteiner*,¹² published in 1930, has a relative frequency of 12.79. This is not so much due to some emotion words having a high frequency, than to the wide array of emotion words present in the play (874 different words).
- The play *D'r Hasestricker*,¹³ published in 1922, has comparatively less emotion words, with a relative frequency of 6.42.

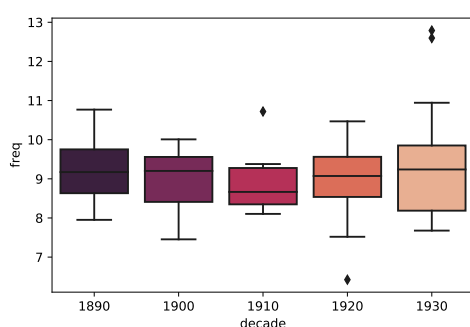


Figure 8: Distribution of the total relative frequency (in %) of words from ELAL in the plays, by decade.

Finally, we compare the distribution of valence scores in ELAL entries and in our theatre plays. Figure 9 displays the kernel density estimates of valence scores in the lexicon and in the corpus of plays. The lexicon is biased towards lower valence entries, as already observed in Figure 6, where the median for valence in ELAL is lower than in the original NRC lexicons. However, in the corpus, this tendency is reversed, with a bias towards higher valence words. This could be due to the genre of the theatre plays, which are predominantly comedies. However, this would require more investigation.

5. Discussion and Perspectives

5.1. Availability of the Lexicon and Research Potential

The emotion lexicon for Alsatian theatre plays is available in two parts on the Nakala open repository.¹⁴ As a by-product, we also obtain multilingual lexicons which could be used in other research settings (online dictionaries, annotation projection) as well as lists of Alsatian graphical variants, which could be used to

¹¹Once all plays will be encoded in TEI, it will be easy to select and deselect plays' structural elements before submitting them to emotion analysis in order to obtain more fine-grained findings about emotion distribution

¹²<https://methal.eu/ui/text/html/wagner-die-greifensteiner/>

¹³<https://methal.eu/ui/text/html/riff-dr-hasestricker/>

¹⁴DOI: 10.34847/nkl.40cex998 (graphical variants) and 10.34847/nkl.39b7617v (scores).

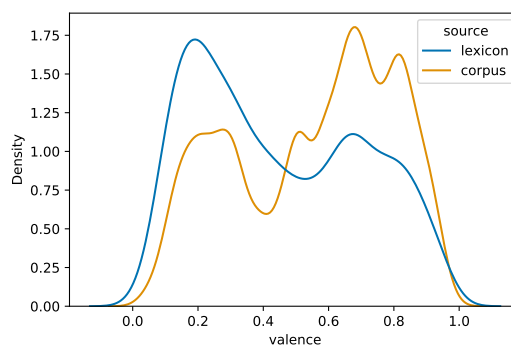


Figure 9: Kernel density estimates of valence scores for entries in ELAL and word forms in the corpus of plays.

study graphical variants in the Alsatian dialects and develop tools to automatically detect them.

5.2. Limitations

As stated in the paper, the final lexicon only represents a fraction of the original NRC lexicons. Due to limited workforce, we were only able to manually correct a part of the data. We saw in Section 4.2.3 that the distribution of scores across the VAD and emotion dimensions was different in ELAL compared to the distribution in the original NRC lexicons: this might impact the analyses performed with ELAL (e.g., bias towards lower valence or higher arousal words).

Moreover, the lexicon has been checked by one person only, and some errors remain. Possible error types are translation errors, semantic drifts, errors in the translations provided in the source NRC lexicons, ambiguities, etc.

5.3. Future Work

We would like to try and adopt a semi-automatic approach to correct the data that has not been manually annotated yet, by taking inspiration from work on multilingual dictionaries (Ács, 2014) or spelling variants filtering (Barteld et al., 2019).

We will also apply the lexicon to investigate theatrical genres with respect to their emotional content, as was done, among others, by Kim et al. (2017) for a corpus of fictional texts in English or Henny-Krahmer (2018) for Spanish American novels; the ongoing project *Emotions in Drama*, which focuses on German theatre (Schmidt et al., 2021), also addresses subgenre-based emotion differences.

6. Acknowledgements

This work has been carried out within the framework of the MeThAl project supported by an IdEx 2020 grant from the Université de Strasbourg. We thank Nathanaël Beiner for the manual correction of the automatically generated emotion networks.

7. Bibliographical References

- Abdaoui, A., Azé, J., Bringay, S., and Poncelet, P. (2017). FEEL: a French Expanded Emotion Lexicon. *Language Resources and Evaluation*, 51(3):833–855, September. Publisher: Springer Verlag.
- Ács, J. (2014). Pivot-based Multilingual Dictionary Building using Wiktionary. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Adolf, P. (2006). *Dictionnaire comparatif multilingue: français-allemand-alsacien-anglais*. Midgard, Strasbourg, France.
- Barteld, F., Biemann, C., and Zinsmeister, H. (2019). Token-based spelling variant detection in Middle Low German texts. *Language Resources and Evaluation*, 53:pages 677–706, February.
- Bernhard, D. (2014). Adding Dialectal Lexicalisations to Linked Open Data Resources: the Example of Alsatian. In *Proceedings of the Workshop on "Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era" at LREC 2014*, pages 23–29, Reykjavík, Islande.
- Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, March.
- Buechel, S., Hellrich, J., and Hahn, U. (2016). Feelings from the Past—Adapting Affective Lexicons for Historical Emotion Analysis. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, pages 54–61.
- Buechel, S., Rücker, S., and Hahn, U. (2020). Learning and Evaluating Emotion Lexicons for 91 Languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1202–1217, Online, July. Association for Computational Linguistics.
- Esuli, A. and Sebastiani, F. (2006). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, May. European Language Resources Association (ELRA).
- Grave, E., Bojanowski, P., Gupta, P., Joulin, A., and Mikolov, T. (2018). Learning word vectors for 157 languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).
- Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In Gaël Varoquaux, et al., editors, *7th Python in Science Conference (SciPy2008)*, pages 11 – 15, Pasadena, CA USA.
- Henny-Krahmer, U. (2018). Exploration of Sentiments and Genre in Spanish American Novels. In Jonathan Girón Palau et al., editors, *13th Annual International Conference of the Alliance of Digital Humanities Organizations, DH 2018, Mexico City, Mexico, June 26-29, 2018, Conference Abstracts*, pages 399–402. Red de Humanidades Digitales.
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95.
- Kim, E. and Klinger, R. (2019). An analysis of emotion communication channels in fan-fiction: Towards emotional storytelling. In *Proceedings of the Second Workshop on Storytelling*, pages 56–64, Florence, Italy, August. Association for Computational Linguistics.
- Kim, E. and Klinger, R. (2021). A Survey on Sentiment and Emotion Analysis for Computational Literary Studies. *Zeitschrift für digitale Geisteswissenschaften*. Publisher: Herzog August Bibliothek Version Number: 2.0.
- Kim, E., Padó, S., and Klinger, R. (2017). Investigating the Relationship between Literary Genres and Emotional Plot Development. In *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 17–26, Vancouver, Canada, August. Association for Computational Linguistics.
- Klinger, R., Kim, E., and Padó, S. (2020). Emotion Analysis for Literary Studies: Corpus Creation and Computational Modelling. In Nils Reiter, et al., editors, *Reflektierte algorithmische Textanalyse*, pages 237–268. De Gruyter, July.
- Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465.
- Mohammad, S. (2011). From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 105–114.
- Mohammad, S. (2018a). Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184, Melbourne, Australia, July. Association for Computational Linguistics.
- Mohammad, S. (2018b). Word Affect Intensities. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).
- Nalisnick, E. T. and Baird, H. S. (2013a). Character-to-character sentiment analysis in Shakespeare's plays. In *Proceedings of the 51st Annual Meeting of*

- the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 479–483.
- Nalisnick, E. T. and Baird, H. S. (2013b). Extracting sentiment networks from Shakespeare’s plays. In *2013 12th International Conference on Document Analysis and Recognition*, pages 758–762. IEEE.
- Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The PageRank Citation Ranking: Bringing Order to the Web. Techreport, Stanford InfoLab, November. Publisher: Stanford InfoLab.
- Philips, L. (2000). The Double Metaphone Search Algorithm. *C/C++ Users Journal*, 18(6):38–43.
- Ramachandran, A. and de Melo, G. (2020). Cross-Lingual Emotion Lexicon Induction using Representation Alignment in Low-Resource Settings. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5879–5890, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Reback, J., jbrockmendel, McKinney, W., den Bossche, J. V., Augspurger, T., Cloud, P., Hawkins, S., gyoung, Sinhrks, Roeschke, M., Klein, A., Petersen, T., Tratner, J., She, C., Ayd, W., Hoefler, P., Naveh, S., Garcia, M., Schendel, J., Hayden, A., Saxton, D., Gorelli, M. E., Shadrach, R., Jancauskas, V., McMaster, A., Li, F., Battiston, P., Seabold, S., Dong, K., and chris b1. (2021). pandas-dev/pandas: Pandas 1.2.5, June.
- Ruiz Fabo, P., Bernhard, D., and Werner, C. (2020). Création d’un corpus FAIR de théâtre en alsacien et normalisation de variétés non-contemporaines. In Thierry Poibeau, et al., editors, *2èmes journées scientifiques du Groupement de Recherche Linguistique Informatique Formelle et de Terrain (LIFT)*, pages 34–43, Montrouge, France. CNRS.
- Ruiz Fabo, P., Werner, C., Bernhard, D., Erhart, P., and Huck, D. (2021). MeThAL : Ressources numériques pour une relecture du théâtre en alsacien. In *10 ans avec CAHIER: Des corpus d’auteurs pour les humanités numériques à leur exploitation numérique (Cahier10)*, Bordeaux, France. Consortium CAHIER.
- Schmidt, T. and Burghardt, M. (2018). An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing. In *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 139–149, Santa Fe, New Mexico, August. Association for Computational Linguistics.
- Schmidt, T., Burghardt, M., and Dennerlein, K. (2018). „Kann man denn auch nicht lachend sehr ernsthaft sein?“ – Zum Einsatz von Sentiment Analyse-Verfahren für die quantitative Untersuchung von Lessings Dramen. In *Book of Abstracts, DHd 2018*.
- Schmidt, T., Dennerlein, K., and Wolff, C. (2021). Emotion Classification in German Plays with Transformer-based Language Models Pretrained on Historical and Contemporary Language. In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 67–79, Punta Cana, Dominican Republic (online), November. Association for Computational Linguistics.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, November.
- Stone, P. J., Dunphy, D., Smith, M. S., and Ogilvie, D. M. (1966). *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press, Cambridge, MA, USA, December.
- TEI Consortium. (2021). TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 4.3.0, August.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.
- Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021. Publisher: The Open Journal.
- Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proc. of HLT-EMNLP-2005*, page 8.

8. Language Resource References

- Abdaoui, A. et al. (2017). *FEEL: French Expanded Emotion Lexicon*. LIRMM.
- Bernhard, D. et. al. (2019). *Annotated Corpus for the Alsatian Dialects*. Zenodo, 2.0.
- Mohammad, S. (2018). *NRC Valence, Arousal, and Dominance (VAD) Lexicon*. National Research Council Canada (NRC), 1.0.
- Mohammad, S. (2020). *The NRC Emotion Intensity Lexicon (NRC-EIL) aka the NRC Affect Intensity Lexicon (NRC-AIL)*. National Research Council Canada (NRC), 1.0.
- Ruiz Fabo, P. et al. (2022). *MeThAL corpus: TEI-encoded corpus of theater plays in Alsatian*. University of Strasbourg, France.